

RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: 10/004,494C
Source: 1FW/b
Date Processed by STIC: 10/3/06

ENTERED



IFW16

RAW SEQUENCE LISTING

DATE: 10/03/2006

PATENT APPLICATION: US/10/004,494C

TIME: 10:22:03

Input Set : E:\10004494 sequence listing.ST25.txt

Output Set: N:\CRF4\10032006\J004494C.raw

```

3 <110> APPLICANT: Cornell Research Foundation
4      Chang, Yung-Fu
6 <120> TITLE OF INVENTION: Ehrlichia canis genes and vaccines
8 <130> FILE REFERENCE: 1258-006 CIP
10 <140> CURRENT APPLICATION NUMBER: 10/004,494C
11 <141> CURRENT FILING DATE: 2001-11-02
13 <150> PRIOR APPLICATION NUMBER: 09/358,322
14 <151> PRIOR FILING DATE: 1999-07-21
16 <160> NUMBER OF SEQ ID NOS: 14
18 <170> SOFTWARE: PatentIn version 3.3
20 <210> SEQ ID NO: 1
21 <211> LENGTH: 5299
22 <212> TYPE: DNA
23 <213> ORGANISM: Ehrlichia canis
26 <220> FEATURE:
27 <221> NAME/KEY: gene
28 <222> LOCATION: (1)..(5299)
30 <400> SEQUENCE: 1
31 gatcaaataa aatgaaacca agaataagaa acactattta tggattaata gcaataatac      60
33 tatctatgat atgttttagtg tacgcttctg taccactata tagtatattt tgtaaagtaa      120
35 cagggttagg aggtacagta agaacaagta atatatcaaa ttctaaaata ggtaacacta      180
37 ttatttaaagt cagattttaat gcagatatata acaaacact gccatggaaa ttctatccag      240
39 aagtatctca tgtatttgta aaaccaggag aacaaaaaatt gattttctac cgcgagaaaa      300
41 atctacttga tgaggacact tcaggaatgg ctgtatataa tgttacacca cataaagtag      360
43 gaaaatattt taataaggta gcttgttttt gtttcaccaa acaaacatta taccctcatc      420
45 aaaaaactat aatgccagta tcatttttta tagatccagc catagaaaca gatcctgaaa      480
47 ctgctgacgt aaaactcatc actctttcat atgtattctt taagtacaaa gaataaactt      540
49 catataccgt acattataaa ctgattaaaa aaaataacta ttaatatga gcaaaataat      600
51 ttatctattc aacagattct tttcaattag agagtattca aaaacactac aactactgct      660
53 tgcaactttc tatcactgat atataaaagt gaaataaatt taaaaaactt tagttttaat      720
55 agaagaattt tattaaaagc tttgaatcaa atttaattac tgatataaaa atactattaa      780
57 acattaacaa tgcttaatta aagtattatt atttacctta atttcataac ctttattaac      840
59 aatttcataa taaaaaactt ttactcttat ttttttatca cttgatatta ttaaataatc      900
61 atataaactc ccaataaac tattgcaagg ttatggtaat gatgaaattt tttacttggt      960
63 ttttcatagt tttcttaaca atagccaatc atgctttatc cttaacatt aaagttacac      1020
65 atgaaaaaatt agataatgga atggaagtat acgtgattcc aaatcatcgc gcaccagcag      1080
67 tcatgcaatc ggtattatac aaagtcggtg gaactgatga tccagtagga tactctggat      1140
69 tagcacattt ttttgaacac ttaatgttta gtggaacaga aaaatttcct aatctcatca      1200
71 gcacacttag taatataggc ggaaatttca atgcaagcac atctcaattt tgtactatat      1260
73 actacgaatt aataccaaaa caatatttat ctcttgcaat ggatattgaa tcagacagaa      1320
75 tgcagaattt taaggttacc gacaaagcat taataagaga acaaaaggta gtcttagaag      1380
77 aaagaaaaat gagagttgaa agccaagcaa aaacatact agaagaagaa atggaaaatg      1440
79 cattttatta caatggatat ggcagaccag tagtaggatg ggaacatgaa attagcaact      1500

```

RAW SEQUENCE LISTING

DATE: 10/03/2006

PATENT APPLICATION: US/10/004,494C

TIME: 10:22:03

Input Set : E:\10004494 sequence listing.ST25.txt

Output Set: N:\CRF4\10032006\J004494C.raw

```

81 acaacaaaga agttgctgaa gcctttcata agctacatta tagtcctaata aatgctatat 1560
83 taattgtaac tggagatgca gatccacaag aagtaatcac acttgcaaaa caatactatg 1620
85 ggaaaatacc atctaataat aagaaacctt caagtcaagt tagggtagaa ccaccgcata 1680
87 aaacaaatat gactttaaca ttaaaagaca gttcagtaga aatcccagaa ctgtttttaa 1740
89 tgtatcaaata accaaatggt attaccaata aaaactacat acttaacatg atggttagcag 1800
91 aaatactcgg tagtggttaa ttcagcctgc tttaaatga tttggtaatt aacaatccaa 1860
93 tagttacatc gataaaaaca gattataatt acttaactga cagcgatgat tagctttcca 1920
95 ttgaagctat acctaaaaac gggatctcta cagaagctgt agaacaagaa attcataaat 1980
97 gtataaataa ttatttagaa aatggaattt cagcagaata tttagaaagt gcaaagtata 2040
99 aagtaaaagc acatttaact tatgcatttg acggactaac tttcatatca ttttttatg 2100
101 gcatgcatct aatactagga gtaccgctat cagaaatcag taatatttac gataccatag 2160
103 acaaagtaag tatccaagat gttaaactccg ctatggaaaa tatctttcaa aacaatataa 2220
105 gattaaccgg gcattttatta cctaattggag aatagttatg agaaacatat tgtgttacac 2280
107 attaatattg attttctttt cattcaatca atatgcaaat gatctcaata ttaacataaa 2340
109 agaagctaca actaaaaata aaatacacta tctatatgtt gaacatcata acctaccaac 2400
111 aatttcctta aaatttgcac tcaagaaagc aggatacgtt tatgatgcct ttgataagca 2460
113 aggacttgca tactttacat caaaaatatt aaacgaagga tcaaaaaaca actatgctct 2520
115 cagttttgca caacaattag aaggcaaagg tatagactta aaatttgata tagacctaga 2580
117 caatttttat atattcaata aaaccttatt agaaaacttt gaagaagccc tagttttact 2640
119 cagtgtttgc atattcaaca ccgtcacaga tcaagaaata ttcattagaa taatagcaga 2700
121 acagattgca catgttaaat catttatattc tgctcctgaa tttatagcta caacagaaat 2760
123 gaatcacgct atattcaaag ggcacccata ttctaacaaa gtttacggga cattaaatac 2820
125 aatcaataat atcaaccagg aagacggtgc attatatata aaaaatagtt ttgacaagga 2880
127 acaaatcggt atcagcgcag caggagatgt agatccaaca cagctatcaa atttactaga 2940
129 taaatatatt ctttccaaat tgccatctgg taataacaaa aataccatac cagatacgac 3000
131 tgtaataaga gaagacacat tatttatgtt acagagagat gtaccacaaa gtgtcataat 3060
133 gtttgctaca gacacagtac catatcacag caaagactat catgcatcaa acttggtcaa 3120
135 tactatgcta ggcggattaa gtctcaattc aatattaatg atagaattaa gagacaagtt 3180
137 aggattaaca taccatagta gcagttcact atctaactg aatcatagta atgtgttatt 3240
139 tggtaacaata ttcactgata ataccacagt aacaaaatgt atatccgtct taacagatat 3300
141 tatagagcac attaaaaagt atggagttga tgaagacact tttgcaattg caaaatctag 3360
143 tattaccaac tcttttattt tatctatgtt aaataacaat aatgttagtg agatattgtt 3420
145 aagcttacaa ttacacgac tagatccgag ttatattaat aaatacaatt cttactacaa 3480
147 agcaataaca atagaagaag taaataaaat tgccaagaaa attttatcta atgaattagt 3540
149 aataattgaa gtaggaaaaa acaataacat aaatggcaaa caaatagatg ctaaaaaaca 3600
151 catacttggg taagtataca gggtattgta tttactacaa gtattctatt aggttgatt 3660
153 aagtaagtat aagtagcttc aatcaaataa aaaaacatta accaaagtgt tagctctacc 3720
155 ggagaagctt attataagct tttaacctgg gataatatga agttttgcta atgttaagca 3780
157 aaaaattagt aatcacaata tcaaattttc tttacaggat tatattgtga cctaccataa 3840
159 caacttatat ttagaaaatg acaacagata cacacatcaa taaattatca ctacaattca 3900
161 attaataaaa caatgagtat ttttacttaa ttatttaatt ttatttttta aaataaaatt 3960
163 acaattttac ttactcaata aaagcagtta tactaccaag tattggatgg tattaatcgg 4020
165 agcaattact acttaatagt atagctgttg acaagccgca atctgcggtt cttgacaaaa 4080
167 taatactaata cagttaaaaa tttgaagtgt ttcaccataa tgggtattat tatgaaagct 4140
169 catagcacia gtatcaggaa ctttcagcct ttgaaaagag ctgctataat cattgcagtg 4200
171 ttaggtttag ctgacttctt gtttgctgct gctgcctgca gtgatcggtt ccaaagattg 4260
173 caattaacaa atccatttgt aatagcagga atgggtggcc ttgcagttct tttagttgct 4320
175 tccttaacag cagcattaag tatatgctta actaaaagta agcaagtcac acaacatgct 4380
177 attagacatc gctttggata cgagtcaagc acttcttctt ctgtactgct tgcaatatca 4440

```

RAW SEQUENCE LISTING

DATE: 10/03/2006

PATENT APPLICATION: US/10/004,494C

TIME: 10:22:03

Input Set : E:\10004494 sequence listing.ST25.txt

Output Set: N:\CRF4\10032006\J004494C.raw

```

179 ataatttctt tattacttgc tgcagcattt tgtggaaaga taatgggtaa tgacaaccca 4500
181 gatctattct ttagcaagat gcaagaactc tccaatccac ttgttggtgc agctattgta 4560
183 gccgtttctg ttttctact ctcattcgtg atgtatgctg caaagaacat tataagtcca 4620
185 gataaacaaa ctcacgttat tatattatct aatcaacaaa ctatagaaga agcaaaagta 4680
187 gatcaaggaa tgaatatttt gtcagcagta ctcccagcag ctggcattga catcatgact 4740
189 atagcttctt gtgacatttt agcagtgcgc agccgggggat cctctcagca tcaatagatt 4800
191 tatgttttag cctgtattca cctttttatt aggtgttgta tcgtttcttt atataagtgt 4860
193 gttatattat ataaaacatc taggagttac agttaatttg tttcatgtgg ttattactct 4920
195 ttgccattat tattactata cctaaaaata taaaagaatc cgccaggttg aatacaggcc 4980
197 aatgtaagtt attgatataa aaatctataa aatcatagac agcaccatat cttattctat 5040
199 ctatgatatt tcctattgac cccccaataa tgattacaag aggtaatcta taatgtggct 5100
201 gtactataaa taagtagcat aaaacacaag taatcaaaat cgagatacta caaaaaacaa 5160
203 cattactata ttcaaagtta tttaatatat caaaactaat tccagcattc cacactgtag 5220
205 taaagcgcaa gaagcttaat atctctatta cacctttatc tcctatcaaa tttactacat 5280
207 accatttact tacctgatc 5299
210 <210> SEQ ID NO: 2
211 <211> LENGTH: 522
212 <212> TYPE: DNA
213 <213> ORGANISM: Ehrlichia canis
216 <220> FEATURE:
217 <221> NAME/KEY: CDS
218 <222> LOCATION: (1)..(522)
219 <223> OTHER INFORMATION: Protein translated from nucleotides 12 through 533
(cytochrome
220 oxidase homolog).
222 <400> SEQUENCE: 2
223 atg aaa cca aga ata aga aac act att tat gga tta ata gca ata ata 48
224 Met Lys Pro Arg Ile Arg Asn Thr Ile Tyr Gly Leu Ile Ala Ile Ile
225 1 5 10 15
227 cta tct atg ata tgt tta gtg tac gct tct gta cca cta tat agt ata 96
228 Leu Ser Met Ile Cys Leu Val Tyr Ala Ser Val Pro Leu Tyr Ser Ile
229 20 25 30
231 ttt tgt aaa gta aca ggt tat gga ggt aca gta aga aca agt aat ata 144
232 Phe Cys Lys Val Thr Gly Tyr Gly Gly Thr Val Arg Thr Ser Asn Ile
233 35 40 45
235 tca aat tct aaa ata ggt aac act att att aaa gtc aga ttt aat gca 192
236 Ser Asn Ser Lys Ile Gly Asn Thr Ile Ile Lys Val Arg Phe Asn Ala
237 50 55 60
239 gat ata cac aaa caa ctg cca tgg aaa ttc tat cca gaa gta tct cat 240
240 Asp Ile His Lys Gln Leu Pro Trp Lys Phe Tyr Pro Glu Val Ser His
241 65 70 75 80
243 gta ttt gta aaa cca gga gaa caa aaa ttg att ttc tac cgc gca gaa 288
244 Val Phe Val Lys Pro Gly Glu Gln Lys Leu Ile Phe Tyr Arg Ala Glu
245 85 90 95
247 aat cta ctt gat gag gac act tca gga atg gct gta tat aat gtt aca 336
248 Asn Leu Leu Asp Glu Asp Thr Ser Gly Met Ala Val Tyr Asn Val Thr
249 100 105 110
251 cca cat aaa gta gga aaa tat ttt aat aag gta gct tgt ttt tgt ttc 384
252 Pro His Lys Val Gly Lys Tyr Phe Asn Lys Val Ala Cys Phe Cys Phe
253 115 120 125

```

RAW SEQUENCE LISTING

DATE: 10/03/2006

PATENT APPLICATION: US/10/004,494C

TIME: 10:22:03

Input Set : E:\10004494 sequence listing.ST25.txt

Output Set: N:\CRF4\10032006\J004494C.raw

```

255 acc aaa caa aca tta tac cct cat caa aaa act ata atg cca gta tca      432
256 Thr Lys Gln Thr Leu Tyr Pro His Gln Lys Thr Ile Met Pro Val Ser
257      130                      135                      140
259 ttt ttt ata gat cca gcc ata gaa aca gat cct gaa act gct gac gta      480
260 Phe Phe Ile Asp Pro Ala Ile Glu Thr Asp Pro Glu Thr Ala Asp Val
261 145                      150                      155                      160
263 aaa ctc atc act ctt tca tat gta ttc ttt aag tac aaa gaa      522
264 Lys Leu Ile Thr Leu Ser Tyr Val Phe Phe Lys Tyr Lys Glu
265                      165                      170
268 <210> SEQ ID NO: 3
269 <211> LENGTH: 174
270 <212> TYPE: PRT
271 <213> ORGANISM: Ehrlichia canis
273 <400> SEQUENCE: 3
275 Met Lys Pro Arg Ile Arg Asn Thr Ile Tyr Gly Leu Ile Ala Ile Ile
276 1                      5                      10                      15
279 Leu Ser Met Ile Cys Leu Val Tyr Ala Ser Val Pro Leu Tyr Ser Ile
280                      20                      25                      30
283 Phe Cys Lys Val Thr Gly Tyr Gly Gly Thr Val Arg Thr Ser Asn Ile
284                      35                      40                      45
287 Ser Asn Ser Lys Ile Gly Asn Thr Ile Ile Lys Val Arg Phe Asn Ala
288      50                      55                      60
291 Asp Ile His Lys Gln Leu Pro Trp Lys Phe Tyr Pro Glu Val Ser His
292 65                      70                      75                      80
295 Val Phe Val Lys Pro Gly Glu Gln Lys Leu Ile Phe Tyr Arg Ala Glu
296                      85                      90                      95
299 Asn Leu Leu Asp Glu Asp Thr Ser Gly Met Ala Val Tyr Asn Val Thr
300                      100                     105                     110
303 Pro His Lys Val Gly Lys Tyr Phe Asn Lys Val Ala Cys Phe Cys Phe
304                      115                     120                     125
307 Thr Lys Gln Thr Leu Tyr Pro His Gln Lys Thr Ile Met Pro Val Ser
308      130                      135                      140
311 Phe Phe Ile Asp Pro Ala Ile Glu Thr Asp Pro Glu Thr Ala Asp Val
312 145                      150                      155                      160
315 Lys Leu Ile Thr Leu Ser Tyr Val Phe Phe Lys Tyr Lys Glu
316                      165                      170
319 <210> SEQ ID NO: 4
320 <211> LENGTH: 1314
321 <212> TYPE: DNA
322 <213> ORGANISM: Ehrlichia canis
325 <220> FEATURE:
326 <221> NAME/KEY: CDS
327 <222> LOCATION: (1)..(1314)
328 <223> OTHER INFORMATION: Protein translated from nucleotides 939 through 2,252
(ProA).
330 <400> SEQUENCE: 4
331 atg atg aaa ttt ttt act tgt ttt ttc ata gtt ttc tta aca ata gcc      48
332 Met Met Lys Phe Phe Thr Cys Phe Phe Ile Val Phe Leu Thr Ile Ala
333 1                      5                      10                      15
335 aat cat gct tta tcc ttt aac att aaa gtt aca cat gaa aaa tta gat      96

```

RAW SEQUENCE LISTING

DATE: 10/03/2006

PATENT APPLICATION: US/10/004,494C

TIME: 10:22:03

Input Set : E:\10004494 sequence listing.ST25.txt

Output Set: N:\CRF4\10032006\J004494C.raw

```

336 Asn His Ala Leu Ser Phe Asn Ile Lys Val Thr His Glu Lys Leu Asp
337          20          25          30
339 aat gga atg gaa gta tac gtg att cca aat cat cgc gca cca gca gtc      144
340 Asn Gly Met Glu Val Tyr Val Ile Pro Asn His Arg Ala Pro Ala Val
341          35          40          45
343 atg cac atg gta tta tac aaa gtc ggt gga act gat gat cca gta gga      192
344 Met His Met Val Leu Tyr Lys Val Gly Gly Thr Asp Asp Pro Val Gly
345          50          55          60
347 tac tct gga tta gca cat ttt ttt gaa cac tta atg ttt agt gga aca      240
348 Tyr Ser Gly Leu Ala His Phe Phe Glu His Leu Met Phe Ser Gly Thr
349          65          70          75          80
351 gaa aaa ttt cct aat ctc atc agc aca ctt agt aat ata ggc gga aat      288
352 Glu Lys Phe Pro Asn Leu Ile Ser Thr Leu Ser Asn Ile Gly Gly Asn
353          85          90          95
355 ttc aat gca agc aca tct caa ttt tgt act ata tac tac gaa tta ata      336
356 Phe Asn Ala Ser Thr Ser Gln Phe Cys Thr Ile Tyr Tyr Glu Leu Ile
357          100          105          110
359 cca aaa caa tat tta tct ctt gca atg gat att gaa tca gac aga atg      384
360 Pro Lys Gln Tyr Leu Ser Leu Ala Met Asp Ile Glu Ser Asp Arg Met
361          115          120          125
363 cag aat ttt aag gtt acc gac aaa gca tta ata aga gaa caa aag gta      432
364 Gln Asn Phe Lys Val Thr Asp Lys Ala Leu Ile Arg Glu Gln Lys Val
365          130          135          140
367 gtc tta gaa gaa aga aaa atg aga gtt gaa agc caa gca aaa aac ata      480
368 Val Leu Glu Glu Arg Lys Met Arg Val Glu Ser Gln Ala Lys Asn Ile
369          145          150          155          160
371 cta gaa gaa gaa atg gaa aat gca ttt tat tac aat gga tat ggc aga      528
372 Leu Glu Glu Glu Met Glu Asn Ala Phe Tyr Tyr Asn Gly Tyr Gly Arg
373          165          170          175
375 cca gta gta gga tgg gaa cat gaa att agc aac tac aac aaa gaa gtt      576
376 Pro Val Val Gly Trp Glu His Glu Ile Ser Asn Tyr Asn Lys Glu Val
377          180          185          190
379 gct gaa gcc ttt cat aag cta cat tat agt cct aat aat gct ata tta      624
380 Ala Glu Ala Phe His Lys Leu His Tyr Ser Pro Asn Asn Ala Ile Leu
381          195          200          205
383 att gta act gga gat gca gat cca caa gaa gta atc aca ctt gca aaa      672
384 Ile Val Thr Gly Asp Ala Asp Pro Gln Glu Val Ile Thr Leu Ala Lys
385          210          215          220
387 caa tac tat ggg aaa ata cca tct aat aat aag aaa cct tca agt caa      720
388 Gln Tyr Tyr Gly Lys Ile Pro Ser Asn Asn Lys Lys Pro Ser Ser Gln
389          225          230          235          240
391 gtt agg gta gaa cca ccg cat aaa aca aat atg act tta aca tta aaa      768
392 Val Arg Val Glu Pro Pro His Lys Thr Asn Met Thr Leu Thr Leu Lys
393          245          250          255
395 gac agt tca gta gaa atc cca gaa ctg ttt tta atg tat caa ata cca      816
396 Asp Ser Ser Val Glu Ile Pro Glu Leu Phe Leu Met Tyr Gln Ile Pro
397          260          265          270
399 aat ggt att acc aat aaa aac tac ata ctt aac atg atg tta gca gaa      864
400 Asn Gly Ile Thr Asn Lys Asn Tyr Ile Leu Asn Met Met Leu Ala Glu

```

RAW SEQUENCE LISTING ERROR SUMMARY DATE: 10/03/2006
PATENT APPLICATION: US/10/004,494C TIME: 10:22:04

Input Set : E:\10004494 sequence listing.ST25.txt
Output Set: N:\CRF4\10032006\J004494C.raw

Invalid <213> Response:

Use of "Artificial" only as "<213> Organism" response is incomplete,
per 1.823(b) of New Sequence Rules. Valid response is Artificial Sequence.

Seq#:12,13,14

VERIFICATION SUMMARY

DATE: 10/03/2006

PATENT APPLICATION: US/10/004,494C

TIME: 10:22:04

Input Set : E:\10004494 sequence listing.ST25.txt

Output Set: N:\CRF4\10032006\J004494C.raw